A Survey on Multimodal Large Language Models for Autonomous Driving

Abstract

With the emergence of Large Language Models (LLMs) and Vision Foundation Models (VFMs), multimodal AI systems benefiting from large models have the potential to equally perceive the real world, make decisions, and control tools as humans. In recent months, LLMs have shown widespread attention in autonomous driving and map systems. Despite its immense potential, there is still a lack of a comprehensive understanding of key challenges, opportunities, and future endeavors to apply in LLM driving systems. In this paper, we present a systematic investigation in this field. We first introduce the background of Multimodal Large Language Models (MLLMs), the multimodal models development using LLMs, and the history of autonomous driving. Then, we overview existing MLLM tools for driving, transportation, and map systems together with existing datasets and benchmarks. Moreover, we summarized the works in The 1st WACV Workshop on Large Language and Vision Models for Autonomous Driving (LLVM-AD), which is the first workshop of its kind regarding LLMs in autonomous driving. To further promote the development of this field, we also discuss several important problems regarding using MLLMs in autonomous driving systems that need to be solved by both academia and industry.